Journal of Translational Medicine

RESEARCH

Open Access

Targeted long-read sequencing facilitates effective carrier screening for complex monogenic diseases including spinal muscular atrophy, α-/β-thalassemia, 21-hydroxylase deficiency, and fragile-X syndrome

Shuyuan Li[1,2,3†], Renyi Hua[1,2,3†], Xu Han[1,2,3†], Yan Xu[1,2,3], Ming Li[1,2,3], Li Gao[1,2,3], Ruiyu Ma[1,2,3], Wanli Meng[4], Aiping Mao[4*], Jian Wang[1,2,3*] and Yanlin Wang[1,2,3*]

Abstract

Background Next-generation sequencing (NGS) has been applied for carrier screening, effectively reducing the incidence of severe diseases. However, some severe, high-prevalent and complex diseases, including spinal muscular atrophy (SMA), α-/β-thalassemia, 21-hydroxylase deficiency (21-OHD), and fragile-X syndrome (FXS), cannot be fully addressed by NGS, resulting in a high residual risk ratio. This study aims to evaluate the clinical utility of a long-read sequencing (LRS) panel for carrier screening of these five complex diseases.

Methods A total of 2926 participants were retrospectively enrolled from International Peace Maternity and Child Health Hospital from Jan 2019 to Dec 2022. All the participants were previously screened for 149 genes correlated to 147 diseases by NGS. The samples were collected and analyzed with the LRS panel targeting the five complex diseases.

Results LRS identified 236 carrier variants, including 54 for SMA, 113 for α-thalassemia, 19 for β-thalassemia, 47 for 21-OHD, and three for FXS. NGS identified only 56.4% (133/236) of the variants detected by LRS. NGS failed to detect three SMA carriers with *SMN1* intragenic variants, while reported 10 false-positive carriers for α-thalassemia (HKαα miscalled as -α3.7). Both 21-OHD and FXS were beyond its detection scope. NGS identified only three of the seven at-risk couples determined by LRS. The total estimated at-risk couple rate for 151 genes in NGS and LRS panels was 1.0996%. SMA, α-/β-thalassemia, 21-OHD, and FXS were among the top 30 high-prevalent diseases and had a combined at-risk couple rate of 0.2433%, accounting for 22.1% of the total ratio. NGS could only identify 22.7% of the at-risk couples for the five diseases in the LRS panel.

†Shuyuan Li, Renyi Hua and Xu Han contributed equally to this work.

*Correspondence:
Aiping Mao
maoaiping@berrygenomics.com
Jian Wang
Labwangjian@shsmu.edu.cn
Yanlin Wang
wyanlin@163.com
Full list of author information is available at the end of the article

Conclusions Comprehensive carrier screening for high-prevalent diseases had higher clinical utility than expanding the list of low-prevalent diseases. Incorporating LRS into the NGS carrier screening strategy would facilitate more effective carrier screening.

Keywords Carrier screening, Thalassemia, SMA, 21-OHD, FXS, Long-read sequencing

Background

Carrier screening is a valuable tool for identifying whether asymptomatic individuals or couples carry gene variants responsible for autosomal recessive (AR) or X-linked (XL) genetic diseases. The primary aim of reproductive carrier screening is to enable informed reproductive decisions. Initiated 50 years ago, carrier screening initially targeted individuals with a family history of specific genetic diseases or populations at higher risk for certain prevalent and severe diseases, such as Ashkenazi Jewish for Tay-Sachs disease [1]. With advancements in next-generation sequencing (NGS) and reduced testing costs, carrier screening has evolved to screen for multiple diseases regardless of ethnicity [1–4]. The American College of Medical Genetics and Genomics (ACMG) has recommended offering carrier screening to all pregnant patients and those planning a pregnancy for diseases with carrier rate of over 1/200, encompassing over a hundred genes [1]. NGS panels targeting dozens to hundreds of genes have been clinically applied for carrier screening, effectively reducing the incidence of severe diseases [3, 5–8].

However, while NGS panels enable expanded list of genes for carrier screening, several high-prevalent diseases, such as spinal muscular atrophy (SMA), α-thalassemia, 21-hydroxylase deficiency (21-OHD), and fragile-X syndrome (FXS), present technical challenges for NGS [9]. The correlated causal genes for these diseases, *SMN1*, *HBA1/2*, *CYP21A2*, and *FMR1*, have high homologous pseudogenes or modifier genes, or low sequence complexity like CGG repeats, making them difficult to be fully addressed by NGS. These diseases contribute substantially to population risk and thus affect the analytical sensitivity of NGS-based carrier screening [9, 10]. Technology advances in experiment and bioinformatics of NGS have enabled the detection of *SMN1* copy number deletion and common variants of α-/β-thalassemia [11–13]. However, residual risks remain, particularly for SMA carrier screening, due to challenges in distinguishing intragenic single-nucleotide variants (SNVs) and indels between *SMN1* and *SMN2*, and in identifying 2 + 0 silent carriers [11, 12]. For thalassemia, complex structural rearrangement between *HBA1* and *HBA2* can lead to inaccurate genotyping results [14]. Additionally, NGS cannot distinguish between functional gene *CYP21A2* and its pseudogene *CYP21A1P* using probe hybridization-based target sequencing, which is the standard method for NGS panel-based carrier screening [15, 16]. For *FMR1*, NGS-based methods rely on computational algorithms to estimate rather than directly determine *FMR1* CGG expanded alleles [17]. The accuracy of these algorithms needs to be further confirmed to be applied clinically [17].

The rapidly developing long-read sequencing (LRS) technologies, such as single-molecule real-time (SMRT) sequencing and nanopore sequencing, are gaining popularity and have been applied in human genetic analysis [18]. LRS-based approaches have been developed for comprehensive genetic analysis of SMA [19, 20], thalassemia [21, 22], 21-OHD [23, 24], and FXS [25]. The analytic sensitivity and clinical utility of these individual LRS assays have been confirmed in genetic diagnosis for patients with clinical features suspected of these diseases, or in genetic screening for small cohorts [19–21, 23, 25–30]. However, these assays have not yet been simultaneously applied to large-scale carrier screening.

The purpose of this study was to evaluate the clinical utility of LRS-based carrier screening for SMA, α-/β-thalassemia, 21-OHD, and FXS, and to assess its benefits for the general population. A LRS panel targeting *SMN1*, *HBA1/2*, *HBB*, *CYP21A2*, and *FMR1* was retrospectively applied to a cohort of 2926 participants who had undergone carrier screening with an NGS panel targeting 149 genes associated with 147 diseases. The carrier rate and at-risk couple rate of the combined 151 genes in the NGS and LRS panels were calculated to demonstrate the cumulative at-risk ratio.

Methods

Study design and participants

In total, 2928 participants were retrospectively enrolled by International Peace Maternity and Child Health Hospital based on the following inclusion criteria: (1) carrier screening with an NGS panel covering 149 genes for 147 diseases had been performed between Jan 2019 to Dec 2022; (2) had no known or reported family history of the diseases included in the NGS or LRS panels; (3) informed written consent was obtained. Two participants were excluded from the study due to unqualified specimen (Fig. 1). This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines.

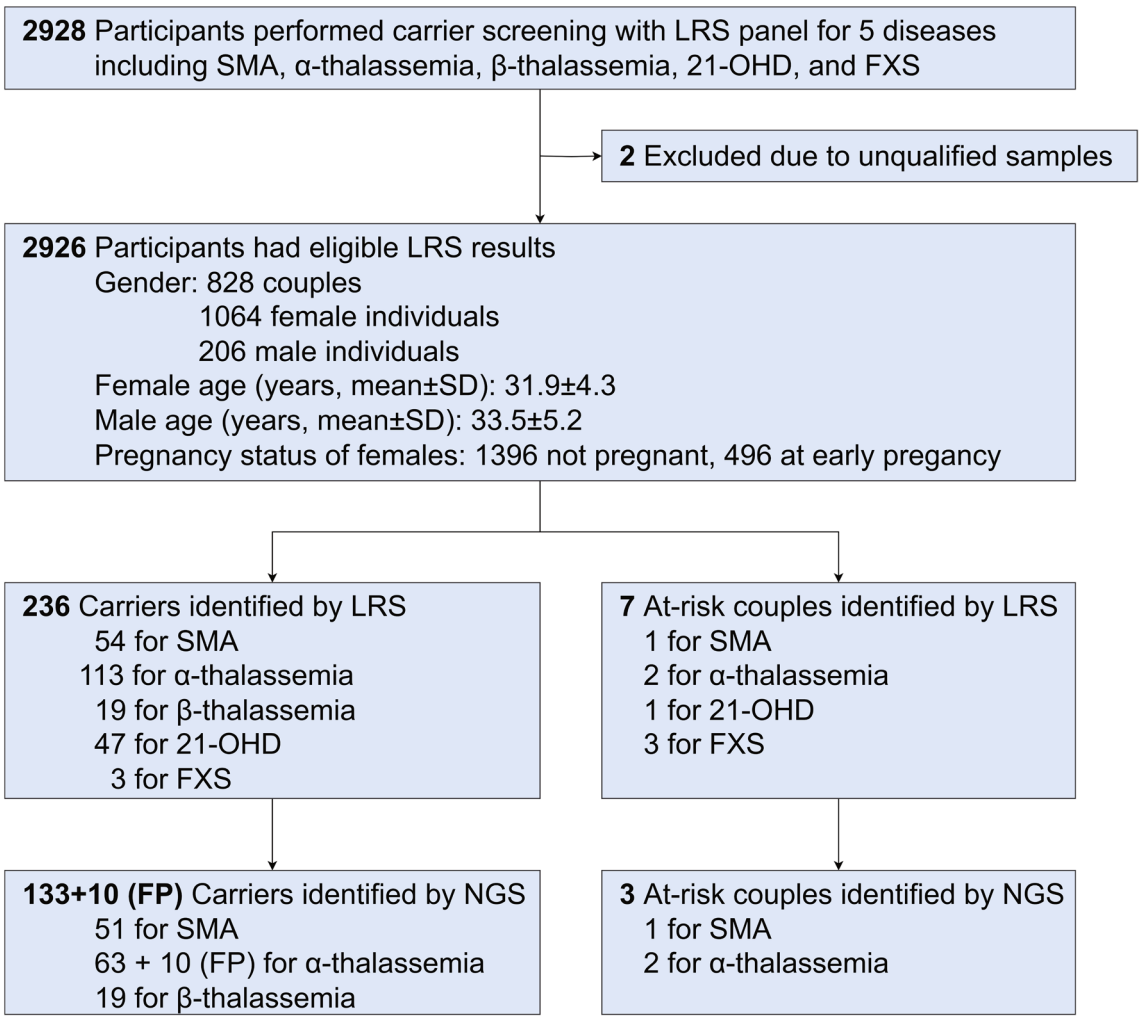

Fig. 1 Study flowchart. FP, false positive

Carrier screening of 147 diseases by NGS panel

A designed panel covering 150 individual genes were used to screen for 147 diseases (Additional file 1: Table S1). For simplicity, α-thalassemia causal genes *HBA1* and *HBA2* were counted as one gene locus because large deletions could encompass both genes. Thus, the NGS panel covered 149 genes, including 139 AR and 10 XL genes. Targeted NGS was performed using the Library Construction and Hybridization Capture Kit and sequencing was performed on the MGISEQ-2000 platform (BGI, Shenzhen, China), as previously described [8]. The pathogenicity of variants was interpreted according to the ACMG guidelines [31], and only pathogenic and likely pathogenic variants were reported. To be mentioned, the hotspot *GJB2*:c.109G > A variant, highly prevalent in East Asia (5.65% carrier rate in this study), was provided in supplemental reports for comprehensive genetic counseling, but excluded from high-risk carrier considerations due to its variable expressivity and incomplete penetrance [32, 33].

Carrier screening of 5 diseases by LRS panel

LRS-based carrier screening for SMA [19, 20], α- and β-thalassemia [21], 21-OHD [23], and FXS [25] were performed for each sample as previously described. Genomic DNA from peripheral blood was extracted with the Blood DNA preparation kit (Genmagbio, Changzhou, China), followed by four multiplex long-range PCR amplification for *SMN1*/*SMN2*, *HBA1*/*HBA2*/*HBB*, *CYP21A2*/*CYP21A1P*, and *FMR1* genes, respectively. Each PCR product was subjected to one-step end-repair and ligation reaction to add unique PacBio barcoded adaptors to get the pre-libraries. The sequencing libraries were prepared using the Sequel II Binding Kit 3.2 (Pacific Biosciences, CA, USA) and sequenced under circular

consensus sequencing (CCS) mode with Sequel II/IIe sequencing platform (Pacific Biosciences, CA, the USA).

Post-sequencing, the raw subreads were processed to get high-quality CCS reads by ccs and debarcoded to individual reaction by lima in the SMRT Link analysis software suite (Pacific Biosciences, CA, the USA). The split CCS reads were then subjected to bioinformatics analysis pipelines for each disease, and the pathogenicity of variants was interpreted according to the ACMG guidelines [31]. Particularly, α-triplicate or quadruplicate were indicated in the reports, as these, when compounded with β-globin gene variants, could lead to intermedia β-thalassemia [28]. For 21-OHD, European Molecular Genetics Quality Network guidelines were also referred to subtype the variants [34]. Samples with salt-wasting or simple-virilizing variants were reported as carriers, while non-classical variants were provided in supplemental reports due to milder and late-onset phenotype [35].

Follow-up

If the spouse of SMA carriers had two or three copies of *SMN1*, the possibility of *SMN1* 2+0 or 3+0 silent carrier was tested with a previously described LRS-based haplotype approach through family trio analysis [20]. The pregnancy status and outcome of at-risk couples for 21-OHD and FXS were retrospectively followed-up and genetic testing of the children and other family members were performed.

Calculation of carrier rate and at-risk couple rate

In total, the NGS and TGS panels covered 151 genes, including 140 AR genes and 11 XL genes, to screen for 149 diseases (Additional file 1: Table S2). For AR genes, the carrier rate was defined as the ratio of carriers among all participants, as no patients with biallelic variants were enrolled in this study. For XL genes, the carrier rate were calculated for females. At-risk couple was defined as follows: both partners carrying a pathogenic or likely pathogenic variant of the same AR gene, or one partner carrying a variant in an XL gene. In the estimation model, the at-risk couple rate for AR genes was calculated as the square of the carrier rate, and for XL genes, it was equal to carrier rate. An exception was made for α-thalassemia: both α^0 (two α-globin genes disrupted) and α^+ (one α-globin gene disrupted) were considered as carrier variants; however, α^+/α^+, which cause only mild α-thalassemia trait, were excluded from the at-risk couple rate calculation for α-thalassemia.

Statistical analysis

Descriptive statistics, such as means and SDs, and graphical representations of the data were derived using SPSS version 24 (IBM Corp). Carrier rate and at-risk couple rate were reported by percentages.

Results

Characteristics of study participants

The cohort included 2926 self-reported Han Chinese participants, with 828 couples, 1064 female individuals, and 206 male individuals (Table 1). The mean (SD) age of female and male participants were 31.9 (4.3) and 33.5 (5.2) years, respectively. Of the 1892 female participants, 1396 were not pregnant, while 496 were at early pregnancy with a mean (SD) gestational age of 10 (1.5) weeks. All samples had genetic screening results that passed quality control for both NGS and LRS panels.

Carriers identified by the LRS panel

The LRS panel targeting five complex genes identified 236 carrier variants, including 54 (1.85%) for SMA, 113 (3.86%) for α-thalassemia, 19 (0.65%) for β-thalassemia, 47 (1.61%) for 21-OHD, and 3 (0.16%) for FXS (Table 1). Among the 54 SMA carriers identified by LRS, NGS identified 51 *SMN1* deletions (94.4%, 51/54), but failed to identify three SNVs/indels. For the 11 SMA carrier families that had trio samples available, LRS haplotype analysis confirmed that none of the 11 spouses were SMA silent carriers (Additional file 1: Fig. S1). For α-thalassemia, LRS detected 16 α^0 variants and 47 α^+ variants, as well as 49 α-triplications and one α-quadruplication. While NGS identified all 63 α^0 and α^+ variants called by LRS, it failed to identified extra α-globin copies, and erroneously called all the 10 functionally normal structural variation HKαα as $-\alpha^{3.7}$ (Additional file 1: Fig. S2), resulting in an error rate of 13.7% (10/73) among positive samples. For β-thalassemia, LRS and NGS had concordant results. 21-OHD and FXS were beyond the detection scope of NGS due to the limitations of short sequencing reads. Consequently, for the five diseases with complex molecular genetics, NGS identified only 56.4% (133/236) of the carriers detected by the LRS panel, meanwhile reported 10 false-positive carriers for α-thalassemia.

SMA, α-thalassemia, and β-thalassemia were screened by both LRS and NGS, and 196 carriers were identified in total, with concordance rates of 94.4% (51/54), 51.2% (63/123), and 100% (19/19) for SMA, α-thalassemia, and β-thalassemia, respectively (Table 1). For SMA, the sensitivity and specificity of NGS were 94.4% (95% confidence interval (CI): 83.7–98.6%) and 100% (95% CI: 99.8–100%). For α-thalassemia, the sensitivity and specificity of NGS were 55.8% (95% CI: 46.1–65.0%) and 99.6% (95% CI: 99.3–99.8%). Both LRS and NGS had 100% of sensitivity and 100% of β-thalassemia.

Table 1 SMA, α-/β-thalassemia, 21-OHD, and FXS carriers identified by LRS and NGS panels from the 2926 participants

Disease	Gene	Variant	Carriers of the variant		Carriers of the gene	
			LRS	NGS	LRS	NGS
SMA	*SMN1*	Deletion	51	51	54	51
		c.22dupA	1	0		
		c.274-1G > A	1	0		
		c.689C > T	1	0		
α-thalassemia	*HBA1*, *HBA2*	*HBA2*:c.342delC (α^+)	1	1	63	63
		HBA2:c.369C > G (α^+)	1	1		
		HBA2:c.427 T > C (α^+)	1	1		
		$--^{SEA}$ (α^0)	15	15		
		$--^{THAI}$ (α^0)	1	1		
		$-\alpha^{27.6}$ (α^+)	1	1		
		$-\alpha^{3.7}$ (α^+)	37	37		
		$-\alpha^{4.2}$ (α^+)	6	6		
		False positive $-\alpha^{3.7a}$	0	10[b]	0	10[b]
		$\alpha\alpha\alpha^{anti3.7}$	29	0	50	0
		$\alpha\alpha\alpha^{anti4.2}$	20	0		
		$\alpha\alpha\alpha\alpha^{anti4.2}$	1	0		
β-thalassemia	*HBB*	c.−78A > G	2	2	19	19
		c.17_18delCT	1	1		
		c.52A > T	2	2		
		c.79G > A	2	2		
		c.92 + 5G > C	1	1		
		c.126_129delCTTT	5	5		
		c.316-197C > T	6	6		
21-OHD	*CYP21A2*	30 kb deletion: *CYP21A1P/A2*-CH-1	5	NA	47	NA
		30 kb deletion: *CYP21A1P/A2*-CH-2	1			
		30 kb deletion: *CYP21A1P/A2*-CH-3	3			
		30 kb deletion: *CYP21A1P/A2*-CH-4	1			
		30 kb deletion: *CYP21A1P/A2*-CH-8	1			
		30 kb deletion: *CYP21A1P/A2*-CH-9	1			
		30 kb deletion: *TNXA/TNXB*-CH-1	2			
		c.124C > T	1			
		c.293-13C/A > G	13			
		c.518T > A	6			
		c.923dupT	1			
		c.955C > T	2			
		c.1069C > T	2			
		c.1306C > T	2			
		c.1451_1452delinsC	2			
		E6_cluster, c.923dupT	1			
		c.923dupT, c.955C > T, c.1069C > T	1			
		c.955C > T, c.1069C > T	2			
FXS	*FMR1*	CGG 23 / 59 repeats	1	NA	3	NA
		CGG 36 / 79 repeats	1			
		CGG 29 / 140 / 168 repeats	1			
Total					236	133 + 10[b]

a, HKαα was misdiagnosed as $-\alpha^{3.7}$ by NGS; b, false-positive carriers identified by NGS; NA, not applicable (not in the detection scope)

At-risk couples identified by the LRS panel

The LRS panel identified seven at-risk couples for the five complex diseases (Table 2). Both LRS and NGS identified the one at-risk family for SMA, and two at-risk families for α-thalassemia. Additionally, LRS identified four at-risk couples that were not in the detection scope of NGS, including three for FXS and one for 21-OHD. For the five complex diseases in the LRS panel, NGS identified only 42.9% (3/7) of the at-risk couples.

Follow up of the couples for the five diseases screened by LRS

The pregnancy status and outcomes of all the 828 couples including 7 at-risk couples for the five diseases identified by LRS were followed up (Table 2). For family CS01 at-risk for SMA, prenatal diagnosis was performed for the first pregnancy and the fetus had homozygous *SMN1* deletion, thus the pregnancy was terminated. The couple was seeking for assisted reproduction technology for a second pregnancy. Both family CS02 and CS03 were at-risk for α-thalassemia. Family CS02 was not pregnant yet at the time of follow-up. Family CS03 did prenatal diagnosis for the fetus, who had normal genotype for α-globin genes and was normally delivered. For the four at-risk families additionally identified by LRS, LRS-based 21-OHD or FXS genetic testing were performed for the children and available family members. In family CS04, both partners carried the heterozygous salt-wasting variant *CYP21A2*:c.293-13C/A > G, making them at-risk of having a child with classical 21-OHD. LRS-based 21-OHD genetic testing revealed that their one-year-old child also had heterozygous *CYP21A2*:c.293-13C/A > G (Fig. 2A). In family CS05, the premutation allele with 59 CGG repeats and 2 AGG interruptions of the mother did not transmit to the daughter, who had two normal *FMR1* alleles (Fig. 2B). In family CS06, the mother had an *FMR1* premutation allele with 79 CGG repeats, and the father had a normal *FMR1* allele with 29 CGG repeats. Their son, diagnosed with developmental delay at two years old, was found by LRS to have a mosaic *FMR1* with both 70 and 415 CGG repeats (Fig. 2C). The couple would plan to perform prenatal diagnosis for a second pregnancy. Family CS07 had a six-month-old son at the time of follow-up who did not show any clinical symptoms. FXS LRS showed that he had an *FMR1* full mutation with mosaic 309 and 509 CGG repeats (Fig. 2D). Genetic testing of the maternal grandparents showed that the grandfather had an *FMR1* premutation allele with 110 CGG repeats, which expanded to mosaic 140/168 repeats during transmission to the mother, and further expanded to full mutation during transmission to the child (Fig. 2D). The allele did not have any AGG interruptions. The couple planned to seek for assisted reproduction technology to have a second child. For the other 821 couples not at-risk of the five diseases, 756 were successfully followed, out of which 538 had one child and 86 had two children at the time of follow up. None of the children had any clinical symptoms of SMA, α-thalassemia, β-thalassemia, 21-OHD, and FXS.

Model of at-risk couple rate for the diseases in LRS and NGS panels

Collectively, the LRS and NGS panels targeted 151 genes, including 140 AR genes and 11 XL genes. Among the AR genes, 30 had a carrier rate over 1/200, including five over 1/50, 10 between 1/50 and 1/100, and 15 between 1/100 and 1/200 (Additional file 1: Table S2). Of note, as the carrier rate of AR genes decreased, the at-risk couple rate decreased dramatically (Fig. 3A). There was a total of 49 AR and XL genes that had at-risk couple rate over 1/100,000 (Table 3). For the 151 genes, the combined at-risk couple rate was 1.0996%. The cumulative ratios for the top 10, 20, 50 and 100 genes were 0.7732%, 0.9710%, 1.0798% and 1.0988%, accounting for 70.3%, 88.3%, 98.2%, 99.9% of the total ratio, respectively (Fig. 3B). FXS and DMD had the highest at-risk couple rate at 0.1586%, while SMA, 21-OHD, α-thalassemia, and β-thalassemia were

Table 2 At-risk couples identified for the 5 genes in the LRS panel

Disease	Gene	Inheritance pattern	Family	Variants	Gender	Panel	At-risk couples
SMA	*SMN1*	AR	CS01	Deletion; Deletion	M; F	LRS, NGS	3
α-Thalassemia	*HBA1*, *HBA2*	AR	CS02	$--^{THAI}$; *HBA2*:c.427T > C	M; F		
			CS03	$--^{SEA}$; $--^{SEA}$	M; F		
21-OHD	*CYP21A2*	AR	CS04	c.293-13C/A > G; c.293-13C/A > G	M; F	LRS	4
FXS	*FMR1*	XL	CS05	CGG 23 / 59 repeats	F		
			CS06	CGG 36 / 79 repeats	F		
			CS07	CGG 29 / 140 / 168 repeats	F		
Total							7

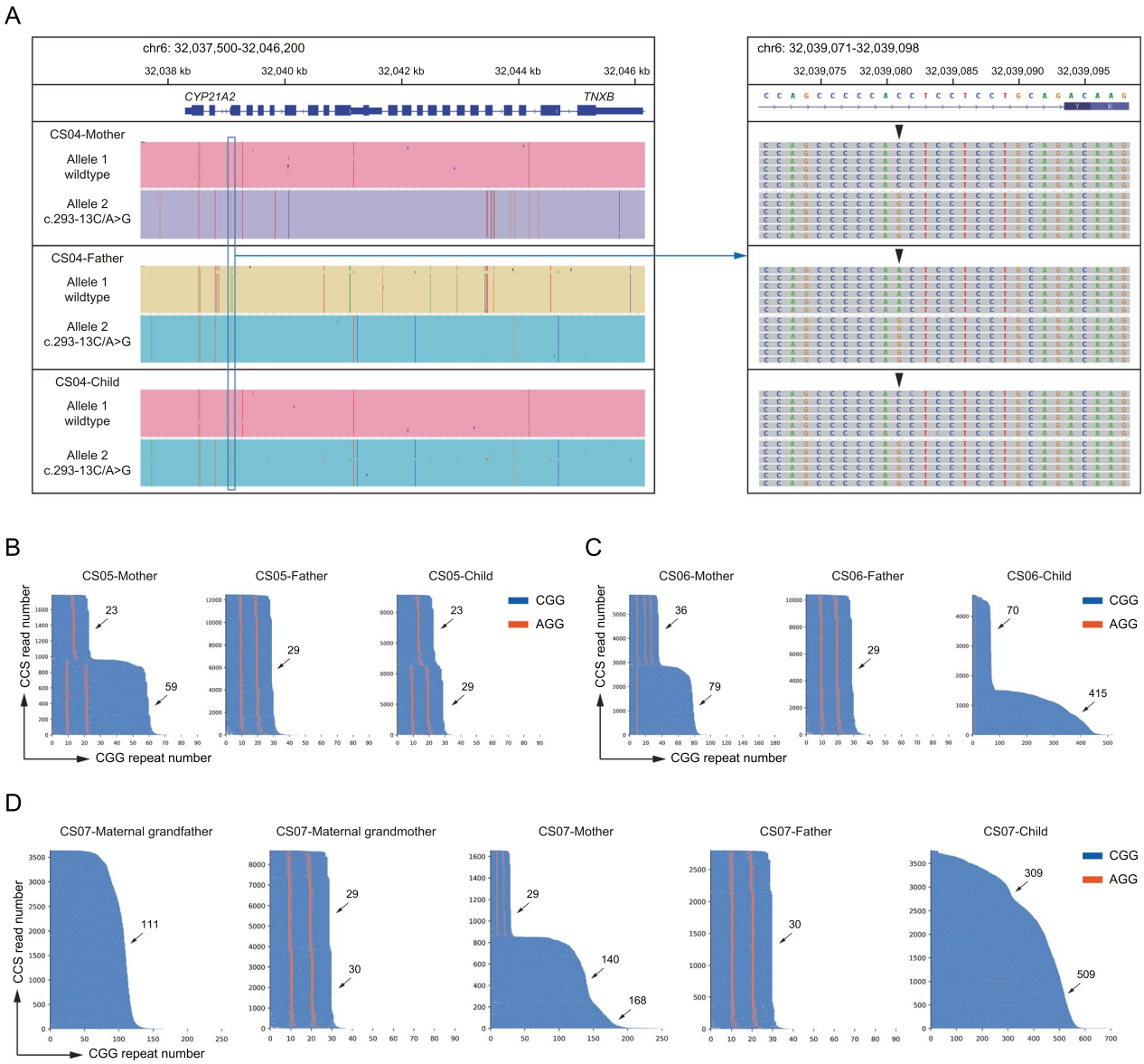

Fig. 2 Follow-up of at-risk couples additionally identified by the LRS panel. **A** Integrative genomics viewer plots displaying the heterozygous *CYP21A2*:c.293-13C/A > G variant of father, mother and child in family CS04. **B**–**D** Waterfall plots showing *FMR1* CGG repeats and AGG interruptions of members in family CS05 (**B**), CS06 (**C**), and CS07 (**D**). The *FMR1* premutation allele in family CS05 was not inherited to the child. The *FMR1* premutation allele in family CS06 was expanded to mosaic pre- and full mutation during transmission. In family CS07, the *FMR1* premutation allele with 110 CGG repeats were expanded to mosaic 140/168 repeats during transmission to the mother, and further expanded to full mutation during transmission to the child

among the top 30, with at-risk couple rate of 0.0341%, 0.0258%, 0.0206%, and 0.0042%, respectively. The combined at-risk couple rate for the five genes in the LRS panel was 0.2433%, accounting for 22.1% of the total rate. NGS could only identify 22.7% of the at-risk couples for the five diseases in the LRS panel. This cumulative study highlighted the importance of performing comprehensive carrier screening for genes with high carrier rates and at-risk couple rates.

Discussion

The development of NGS techniques have enabled carrier screening with expanded panels containing hundreds of diseases. However, some high-prevalent and complex diseases, including SMA, α-/β-thalassemia, 21-OHD, and FXS, could not be effectively screened by NGS due to limited read length. This retrospective study demonstrated that comprehensively addressing these complex diseases had much higher clinical yield

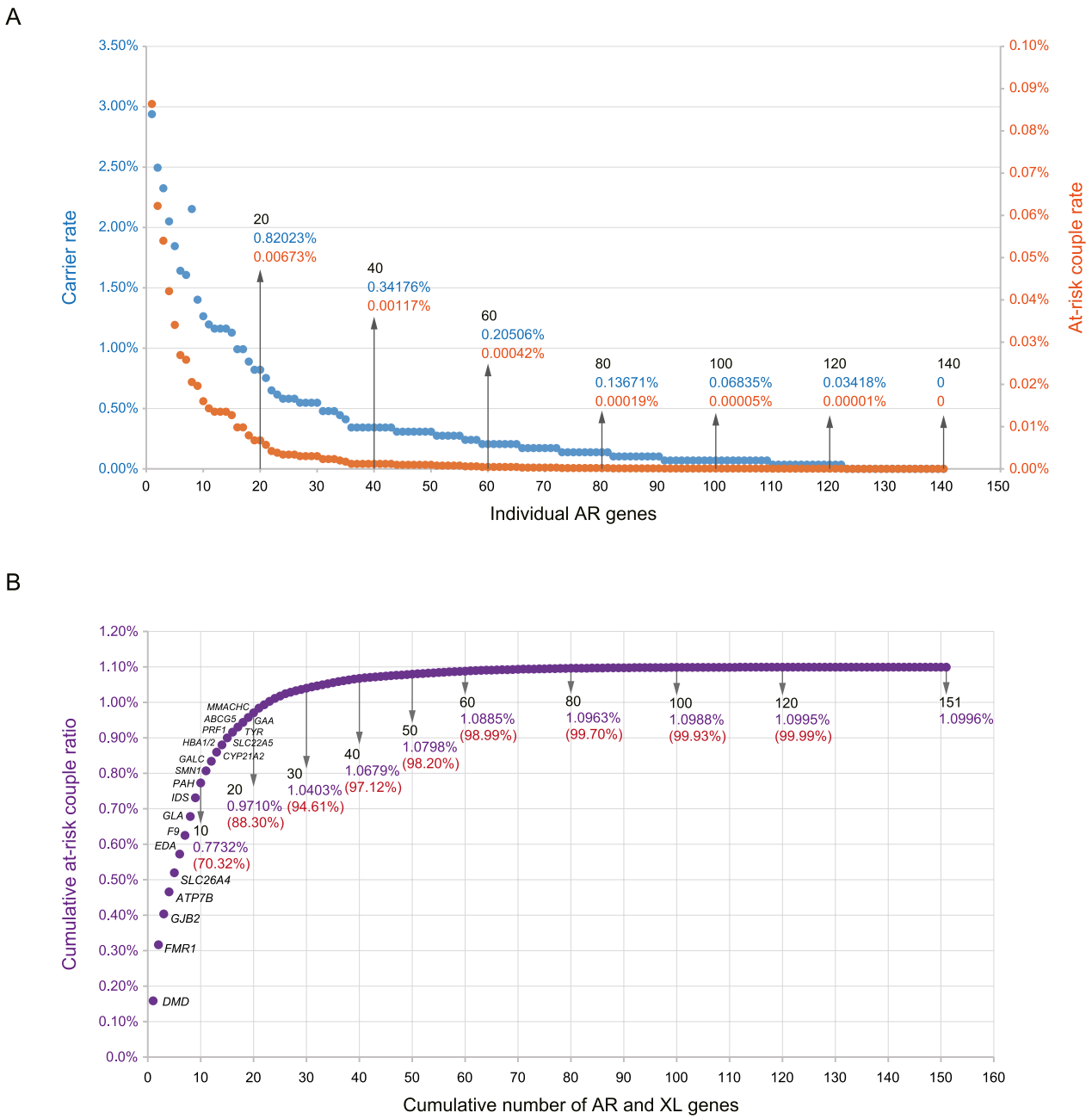

Fig. 3 Carrier rate and at-risk couple rate for the genes in the LRS and NGS panels. **A** Carrier rate and at-risk couple rate for the 140 AR genes. **B** Cumulative at-risk couple rate for all the 151 genes, with top 20 genes labeled in the plot

compared to add more genes with low carrier rate. An LRS panel targeting these five diseases was applied for carrier screening in a cohort of 2926 participants who had also undergone NGS-based carrier screening. Overall, NGS identified only 56.4% (133/236) of the carriers identified by LRS, and 42.9% (3/7) of the at-risk couples in the cohort identified by LRS.

Compared to NGS, LRS additionally identified three SMA carriers with SNVs/indels and demonstrated its ability to determine whether a participant with normal *SMN1* copy number was a silent carrier. LRS could successfully identify the structural rearrangements caused by homologous recombination in *HBA1/2*. It accurately called all the 10 samples with functionally normal

Table 3 List of diseases and genes with estimated at-risk couple rate over 1/100,000

No	Diseases	Gene	Inheritance pattern	Carriers	Carrier rate %	At-risk couple rate %
1	Duchenne Muscular Dystrophy	*DMD*	XL	3	0.1586	0.1586
2	Fragile-X syndrome	*FMR1*	XL	3	0.1586	0.1586
3	Autosomal Recessive Deafness 1A	*GJB2*	AR	86	2.9392	0.0864
4	Wilson Disease	*ATP7B*	AR	73	2.4949	0.0622
5	Autosomal Recessive Deafness 4, with Enlarged Vestibular Aqueduct	*SLC26A4*	AR	68	2.3240	0.0540
6	X-Linked Hypohidrotic Ectodermal Dysplasia	*EDA*	XL	1	0.0529	0.0529
7	Hemophilia B	*F9*	XL	1	0.0529	0.0529
8	Fabry Disease	*GLA*	XL	1	0.0529	0.0529
9	Mucopolysaccharidosis II	*IDS*	XL	1	0.0529	0.0529
10	Phenylketonuria	*PAH*	AR	60	2.0506	0.0420
11	Spinal Muscular Atrophy	*SMN1*	AR	54	1.8455	0.0341
12	Krabbe Disease	*GALC*	AR	48	1.6405	0.0269
13	21-hydroxylase deficiency	*CYP21A2*	AR	47	1.6063	0.0258
14	α-thalassemia	*HBA1, HBA2*	AR	63	2.1531	0.0206
15	Primary Carnitine Deficiency	*SLC22A5*	AR	41	1.4012	0.0196
16	Hemophagocytic lymphohistiocytosis, familial, 2	*PRF1*	AR	37	1.2645	0.0160
17	Oculocutaneous Albinism Type IA and IB	*TYR*	AR	35	1.1962	0.0143
18	Sitosterolemia 2	*ABCG5*	AR	34	1.1620	0.0135
19	Glycogen Storage Disease Type II	*GAA*	AR	34	1.1620	0.0135
20	Methylmalonic Aciduria and Homocystinuria cblC type	*MMACHC*	AR	34	1.1620	0.0135
21	Glutaric acidemia IIC	*ETFDH*	AR	33	1.1278	0.0127
22	MUT-Related Methylmalonic Acidemia	*MMUT*	AR	29	0.9911	0.0098
23	Oculocutaneous Albinism Type 2	*OCA2*	AR	29	0.9911	0.0098
24	Cystic Fibrosis	*CFTR*	AR	26	0.8886	0.0079
25	Glycogen Storage Disease Type Ia	*G6PC*	AR	24	0.8202	0.0067
26	Hemophagocytic lymphohistiocytosis, familial, 3	*UNC13D*	AR	24	0.8202	0.0067
27	β-hemoglobinopathies (including β-thalassemia and sickle cell disease)	*HBB*	AR	19	0.6494	0.0042
28	Smith-Lemli-Opitz syndrome	*DHCR7*	AR	19	0.6494	0.0042
29	Short-chain acyl-CoA dehydrogenase deficiency (SCAD deficiency)	*ACADS*	AR	18	0.6152	0.0038
30	Hypophosphatasia, infantile and childhood	*ALPL*	AR	17	0.5810	0.0034
31	Limb-girdle muscular dystrophy type 2, subtypes 2A	*CAPN3*	AR	17	0.5810	0.0034
32	Megalencephalic Leukoencephalopathy with Subcortical Cysts 1	*MLC1*	AR	17	0.5810	0.0034
33	Joubert syndrome 5, Meckel Syndrome 4	*CEP290*	AR	16	0.5468	0.0030
34	Limb-Girdle Muscular Dystrophy type 2B	*DYSF*	AR	16	0.5468	0.0030
35	Nephrotic syndrome, type 1	*NPHS1*	AR	16	0.5468	0.0030
36	Oculocutaneous Albinism Type 4	*SLC45A2*	AR	16	0.5468	0.0030
37	Medium-chain acyl-CoA dehydrogenase deficiency	*ACADM*	AR	14	0.4785	0.0023
38	Ellis-van Creveld Syndrome	*EVC2*	AR	14	0.4785	0.0023
39	TMEM67-related disorders	*TMEM67*	AR	14	0.4785	0.0023
40	Alport syndrome 2, autosomal recessive	*COL4A4*	AR	13	0.4443	0.0020
41	Niemann-Pick Disease Type C1 and D	*NPC1*	AR	12	0.4101	0.0017
42	Very long-chain acyl-CoA dehydrogenase (VLCAD) deficiency	*ACADVL*	AR	10	0.3418	0.0012
43	Joubert syndrome 3	*AHI1*	AR	10	0.3418	0.0012
44	Biotinidase Deficiency	*BTD*	AR	10	0.3418	0.0012
45	Autosomal Recessive Epidermolysis Bullosa Dystrophica	*COL7A1*	AR	10	0.3418	0.0012
46	Joubert Syndrome 17	*CPLANE1*	AR	10	0.3418	0.0012
47	Hermansky-Pudlak Syndrome 3	*HPS3*	AR	10	0.3418	0.0012
48	Propionicacidemia	*PCCB*	AR	10	0.3418	0.0012

Table 3 (continued)

No	Diseases	Gene	Inheritance pattern	Carriers	Carrier rate %	At-risk couple rate %
49	Hyperphenylalaninemia, BH4-deficient, A	*PTS*	AR	10	0.3418	0.0012

variant HKαα, which NGS miscalled as -$\alpha^{3.7}$. Moreover, 50 participants with α-triplicate or quadruplicate were identified by LRS. β-thalassemia intermedia caused by heterozygous β-globin gene variant co-inherited with α-triplicate or quadruplicate could be clinically heterogeneous, ranging from asymptotic to transfusion dependent thalassemia [36]. However, there are higher needs for blood transfusion in women, especially during pregnancy or delivery [36]. This is particularly important for regions with high-prevalence of β-thalassemia.

Particularly, LRS identified four at-risk couples for 21-OHD and FXS. The LRS genetic testing results and associated pregnancy risks were informed to all the four families for future reproductive options. To be mentioned, carrier screening for FXS mainly focuses on female premutation carriers, as the CGG premutation alleles are transmitted from fathers to daughters in a relatively stable manner and can also contract, though in very rare cases they can also expand to full mutation [37]. In family CS07, the grandfather's 110 CGG repeats expanded to mosaic 140/168 repeats during transmission to the mother and further expanded to a full mutation in the next generation, highlighting the importance of screening for male premutation carriers as well. These results demonstrated the high analytic sensitivity and clinical utility of LRS-based carrier screening for diseases with challenging sequence features.

The cumulative carrier rate was 52.2032%, and at-risk couple rate was 1.0996%. For the 140 AR genes, only 30 had carrier rate over 1/200. In total, 49 AR and XL genes had an at-risk couple rate of over 1/100,000. The cumulative at-risk ratio of these 49 genes was 1.0789%, accounting for 98.1% of the total ratio. The five diseases in LRS panel had a combined at-risk couple rate of 0.2433%, accounting for 22.1% of the total ratio, while NGS could only identify 22.7% of these at-risk couples. This demonstrated that instead of adding more genes with low carrier rate, the more effective approach should be increase the detection rate of genes with high carrier rate.

With the same carrier rate, XL diseases have much higher at-risk couple rate than AR diseases, thus it is particularly important to perform carrier screening for XL genes. XL gene *F8* also has complicated molecular genetics due to inversions caused by homologous recombination. Approximately 40–50% of *F8* disease-causing variants were inversions of intron 22 and intron 1, which were beyond the detection scope of NGS but could be addressed by LRS [38]. Including these molecular challenging genes would further facilitate effective carrier screening.

Though the main purpose of carrier screening is for the management of reproductive risk, incidental molecular diagnosis could identify variants in genes that may confer increased risk for adult-onset conditions [39]. NGS identified that two individuals had two variants of *GJB2*. One had homozygous *GJB2*:c.109G > A, and the other had both *GJB2*:c.109G > A and *GJB2*:c.235delC. LRS identified that one individual had *CYP21A2* genotype (c.[−126C > T, c.−113G > A]/E6cluster, c.923dup) correlated to non-classical 21-OHD. Additionally, three females carried premutation alleles with CGG repeat of 59, 79, and 140/168, respectively. These individuals might be at increased risk of fragile-X-associated conditions.

While SMA, α-thalassemia, and β-thalassemia could be screened by both LRS and NGS (though NGS had lower assay performance), 21-OHD and FXS could only be effectively screened by LRS. Currently, multiplex ligation-dependent probe amplification combined with Sanger sequencing is employed for genetic diagnosis of 21-OHD [29]. However, this approach is methodologically complicated and not feasible for large-scale genetic carrier screening. While PCR-capillary electrophoresis method could be used for FXS carrier screening [40], it is difficult to screen for a panel of genes with the same platform. In the present study, the LRS sequencing libraries for SMA, α-/β-thalassemia, 21-OHD, and FXS were prepared separately and pooled for sequencing, with cost of approximately 100 US dollars per sample, and turnaround time of 8 days. Developing assays that can perform sequencing library for all the five diseases simultaneously could possibly decrease the cost per sample to about 50 US dollars. With the increase of LRS throughput and decrease of sequencing cost, there would be high financial and operational feasibility of LRS application in clinical setting for screening both complex and non-complex monogenic diseases.

There are several limitations of this study. First, this is a retrospective, single-center study conducted in Shanghai, and the carrier rate of some genes may vary geographically. For instance, this cohort has much lower carrier rate of α- and β-thalassemia compared to southern cities in China [21]. Second, some XL genes

only have one carrier identified each due to limited number of participants. Future prospective, multi-center, large-scale clinical studies are warranted to get more accurate calculation of carrier rate and at-risk couple rate in different areas of Chinese population.

Conclusions

LRS has demonstrated high clinical utility for carrier screening of molecular challenging diseases, including SMA, α-/β-thalassemia, 21-OHD, and FXS, which has much better clinical yield than expanding the NGS panels to include more low-prevalent diseases. Incorporating LRS panel into the current NGS panel-based carrier screening program will facilitate more effective carrier screening and assist for informed reproduction decision.

Abbreviations

LRS	Long-read sequencing
SMA	Spinal muscular atrophy
21-OHD	21-Hydroxylase deficiency
FXS	Fragile-X syndrome
NGS	Next-generation sequencing
AR	Autosomal recessive
XL	X-linked
ACMG	The American College of Medical Genetics and Genomics
SNV	Single-nucleotide variant
SMRT	Single-molecule real-time
CNV	Copy-number variation
CCS	Circular consensus sequencing

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-025-06345-1.

Additional file 1

Additional file 2

Additional file 3

Additional file 4

Acknowledgements

We thank all the patients for their participation in this study.

Author contributions

Conceptualization, YW, JW, SL; sampling and clinical follow up, RH, XH; sequencing and data analysis, AM, WM; data curation, YX, ML, LG, RM; project administration and funding acquisition, YW, SL; writing and editing, YW, JW, SL, RH, XH and AM. All authors have read and approved the final manuscript.

Funding

This study was supported by the Shanghai Municipal Science and Technology Commission (22Y11902300, 22Y11906700), Clinical Research Plan of IPMCH (IPMCH2022CR1-02), the National Natural Science Foundation of China (No. 81971401, 81871136), the National Key Technology Research and Development Program of the Ministry of Science and Technology (2023YFC2705901), and the Shanghai Municipal Key Clinical Specialty.

Availability of data and materials

The data would be available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Approval was obtained from the Ethics Committee of International Peace Maternity and Child Health Hospital (#GKLW-A-2024-027-01). Informed consent was obtained from all participants or their legal guardians involved in the study.

Consent for publication

Informed consent was obtained from all participants or their legal guardians for publication of the data.

Competing interests

A.M. and W.M. are employees of Berry Genomics Corporation. No other disclosures are reported.

Author details

[1]International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University, 910 Hengshan Road, Shanghai 200030, China. [2]Shanghai Key Laboratory of Embryo Original Diseases, Shanghai 200030, China. [3]Institute of Birth Defects and Rare Diseases, School of Medicine, Shanghai Jiao Tong University, Shanghai 200030, China. [4]Berry Genomics Corporation, Beijing 102200, China.

Received: 22 July 2024 Accepted: 4 March 2025
Published online: 11 March 2025

References

1. Gregg AR, Aarabi M, Klugman S, Leach NT, Bashford MT, Goldwaser T, Chen E, Sparks TN, Reddi HV, Rajkovic A, Dungan JS. Screening for autosomal recessive and X-linked conditions during pregnancy and preconception: a practice resource of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2021;23:1793–806.
2. Westemeyer M, Saucier J, Wallace J, Prins SA, Shetty A, Malhotra M, Demko ZP, Eng CM, Weckstein L, Boostanfar R, et al. Clinical experience with carrier screening in a general population: support for a comprehensive pan-ethnic approach. Genet Med. 2020;22:1320–8.
3. Haque IS, Lazarin GA, Kang HP, Evans EA, Goldberg JD, Wapner RJ. Modeled fetal risk of genetic diseases identified by expanded carrier screening. JAMA. 2016;316:734–42.
4. Guo MH, Gregg AR. Estimating yields of prenatal carrier screening and implications for design of expanded carrier screening panels. Genet Med. 2019;21:1940–7.
5. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, Langley RJ, Zhang L, Lee CC, Schilkey FD, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci Transl Med. 2011;3:65ra64.
6. Beauchamp KA, Johansen Taber KA, Muzzey D. Clinical impact and cost-effectiveness of a 176-condition expanded carrier screen. Genet Med. 2019;21:1948–57.
7. Zhao S, Xiang J, Fan C, Asan, Shang X, Zhang X, Chen Y, Zhu B, Cai W, Chen S, et al. Pilot study of expanded carrier screening for 11 recessive diseases in China: results from 10,476 ethnically diverse couples. Eur J Hum Genet. 2019;27:254–62.
8. Zhang X, Chen Q, Li J, Luo X, Luo J, Li J, Zeng Z, Wu Y, Zhang H, Dong Y. The effectiveness of expanded carrier screening based on next-generation sequencing for severe monogenic genetic diseases. Hum Genomics. 2024;18:9.
9. Beauchamp KA, Muzzey D, Wong KK, Hogan GJ, Karimi K, Candille SI, Mehta N, Mar-Heyming R, Kaseniit KE, Kang HP, et al. Systematic design and comparison of expanded carrier screening panels. Genet Med. 2018;20:55–63.
10. Punj S, Akkari Y, Huang J, Yang F, Creason A, Pak C, Potter A, Dorschner MO, Nickerson DA, Robertson PD, et al. Preconception carrier screening by genome sequencing: results from the clinical laboratory. Am J Hum Genet. 2018;102:1078–89.
11. Feng Y, Ge X, Meng L, Scull J, Li J, Tian X, Zhang T, Jin W, Cheng H, Wang X, et al. The next generation of population-based spinal muscular atrophy

carrier screening: comprehensive pan-ethnic SMN1 copy-number and sequence variant analysis by massively parallel sequencing. Genet Med. 2017;19:936–44.
12. Zhao S, Wang W, Wang Y, Han R, Fan C, Ni P, Guo F, Zeng F, Yang Q, Yang Y, et al. NGS-based spinal muscular atrophy carrier screening of 10,585 diverse couples in China: a pan-ethnic study. Eur J Hum Genet. 2021;29:194–204.
13. He J, Song W, Yang J, Lu S, Yuan Y, Guo J, Zhang J, Ye K, Yang F, Long F, et al. Next-generation sequencing improves thalassemia carrier screening among premarital adults in a high prevalence population: the Dai nationality, China. Genet Med. 2017;19:1022–31.
14. Huang R, Liu Y, Xu J, Lin D, Mao A, Yang L, Zhong G, Wang H, Xu R, Chen Y, Zhou Q. Back-to-back comparison of third-generation sequencing and next-generation sequencing in carrier screening of Thalassemia. Arch Pathol Lab Med. 2023;148:797.
15. Gangodkar P, Khadilkar V, Raghupathy P, Kumar R, Dayal AA, Dayal D, Ayyavoo A, Godbole T, Jahagirdar R, Bhat K, et al. Clinical application of a novel next generation sequencing assay for CYP21A2 gene in 310 cases of 21- hydroxylase congenital adrenal hyperplasia from India. Endocrine. 2021;71:189–98.
16. Wang W, Han R, Yang Z, Zheng S, Li H, Wan Z, Qi Y, Sun S, Ye L, Ning G. Targeted gene panel sequencing for molecular diagnosis of congenital adrenal hyperplasia. J Steroid Biochem Mol Biol. 2021;211: 105899.
17. Bahlo M, Bennett MF, Degorski P, Tankard RM, Delatycki MB, Lockhart PJ. Recent advances in the detection of repeat expansions with short-read next-generation sequencing. F1000Res. 2018;7:736.
18. Oehler JB, Wright H, Stark Z, Mallett AJ, Schmitz U. The application of long-read sequencing in clinical settings. Hum Genomics. 2023;17:73.
19. Li S, Han X, Xu Y, Chang C, Gao L, Li J, Lu Y, Mao A, Wang Y. Comprehensive analysis of spinal muscular atrophy: SMN1 copy number, intragenic mutation, and 2 + 0 carrier analysis by third-generation sequencing. J Mol Diagn. 2022;24:1009–20.
20. Li S, Han X, Zhang L, Xu Y, Chang C, Gao L, Zhan J, Hua R, Mao A, Wang Y. An effective and universal long-read sequencing-based approach for SMN1 2 + 0 carrier screening through family trio analysis. Clin Chem. 2023;69:1295–306.
21. Liang Q, Gu W, Chen P, Li Y, Liu Y, Tian M, Zhou Q, Qi H, Zhang Y, He J, et al. A more universal approach to comprehensive analysis of thalassemia alleles (CATSA). J Mol Diagn. 2021;23:1195–204.
22. Huang W, Qu S, Qin Q, Yang X, Han W, Lai Y, Chen J, Zhou S, Yang X, Zhou W. Nanopore third-generation sequencing for comprehensive analysis of hemoglobinopathy variants. Clin Chem. 2023;69:1062–71.
23. Liu Y, Chen M, Liu J, Mao A, Teng Y, Yan H, Zhu H, Li Z, Liang D, Wu L. Comprehensive analysis of congenital adrenal hyperplasia using long-read sequencing. Clin Chem. 2022;68:927–39.
24. Tantirukdham N, Sahakitrungruang T, Chaisiwamongkol R, Pongpanich M, Srichomthong C, Assawapitaksakul A, Buasong A, Tongkobpetch S, Yeetong P, Shotelersuk V. Long-read amplicon sequencing of the CYP21A2 in 48 Thai patients with steroid 21-hydroxylase deficiency. J Clin Endocrinol Metab. 2022;107:1939–47.
25. Liang Q, Liu Y, Liu Y, Duan R, Meng W, Zhan J, Xia J, Mao A, Liang D, Wu L. Comprehensive analysis of fragile X syndrome: full characterization of the FMR1 locus by long-read sequencing. Clin Chem. 2022;68:1529–40.
26. Hou F, Mao A, Shan S, Li Y, Meng W, Zhan J, Nie W, Jin H. Evaluating the clinical utility of a long-read sequencing-based approach in genetic testing of fragile-X syndrome. Clin Chim Acta. 2023;551: 117614.
27. Bai J, Qu Y, Huang W, Meng W, Zhan J, Wang H, Hou W, Jin Y, Mao A, Song F. A high-fidelity long-read sequencing-based approach enables accurate and effective genetic diagnosis of spinal muscular atrophy. Clin Chim Acta. 2024;553: 117743.
28. Liang Q, He J, Li Q, Zhou Y, Liu Y, Li Y, Tang L, Huang S, Li R, Zeng F, et al. Evaluating the clinical utility of a long-read sequencing-based approach in prenatal diagnosis of Thalassemia. Clin Chem. 2023;69:239–50.
29. Yuan D, Cai R, Mao A, Tan J, Zhong Q, Zeng D, Tang N, Wei X, Huang J, Zhang Y, et al. Improved genetic characterization of congenital adrenal hyperplasia by long-read sequencing compared with multiplex ligation-dependent probe amplification plus sanger sequencing. J Mol Diagn. 2024;26:770.
30. Zhang R, Cui D, Song C, Ma X, Cai N, Zhang Y, Feng M, Cao Y, Chen L, Qiang R. Evaluating the efficacy of a long-read sequencing-based approach in the clinical diagnosis of neonatal congenital adrenocortical hyperplasia. Clin Chim Acta. 2024;555: 117820.
31. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405–24.
32. Chen T, Fan C, Huang Y, Feng J, Zhang Y, Miao J, Wang X, Li Y, Huang C, Jin W, et al. Genomic sequencing as a first-tier screening test and outcomes of newborn screening. JAMA Netw Open. 2023;6: e2331162.
33. Zhu Y, Hu L, Yang L, Wang L, Lu Y, Dong X, Xiao T, Xu Z, Wu B, Zhou W. Association between expanded genomic sequencing combined with hearing screening and detection of hearing loss among newborns in a neonatal intensive care unit. JAMA Netw Open. 2022;5: e2220986.
34. Baumgartner-Parzer S, Witsch-Baumgartner M, Hoeppner W. EMQN best practice guidelines for molecular genetic testing and reporting of 21-hydroxylase deficiency. Eur J Hum Genet. 2020;28:1341–67.
35. Auer MK, Nordenström A, Lajic S, Reisch N. Congenital adrenal hyperplasia. Lancet. 2023;401:227–44.
36. Bonello-Palot N, Benoit A, Agouti I, Hamouda I, Brousse V, Badens C. Transfusion requirements and complication rate in β-thalassemia intermedia due to heterozygous β-globin gene mutation and triplicated α-globin genes. Eur J Haematol. 2023;111:742–7.
37. Spector E, Behlmann A, Kronquist K, Rose NC, Lyon E, Reddi HV. Laboratory testing for fragile X, 2021 revision: a technical standard of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2021;23:799–812.
38. Liu Y, Li D, Yu D, Liang Q, Chen G, Li F, Gao L, Li Z, Xie T, Wu L, et al. Comprehensive analysis of hemophilia A (CAHEA): towards full characterization of the f8 gene variants by long-read sequencing. Thromb Haemost. 2023;123:1151–64.
39. Reiner J, Rosenblum LS, Xin W, Zhou Z, Zhu H, Leach N. Incidental molecular diagnoses and heterozygous risk alleles in a carrier screening cohort. Genet Med. 2023;25: 100317.
40. Guo Q, Chang YY, Huang CH, Hsiao YS, Hsiao YC, Chiu IF, Zhou Y, Zhang H, Ko TM. Population-based carrier screening and prenatal diagnosis of fragile X syndrome in East Asian populations. J Genet Genomics. 2021;48:1104–10.

Publisher's Note